

Accessibility-based reranking in multimedia search engines

Ilias Kalamaras¹  · Nikolaos Dimitriou² ·
Anastasios Drosou² · Dimitrios Tzovaras²

Received: 7 December 2015 / Revised: 9 June 2016 / Accepted: 17 August 2016 /

Published online: 10 September 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Traditional multimedia search engines retrieve results based mostly on the query submitted by the user, or using a log of previous searches to provide personalized results, while not considering the accessibility of the results for users with vision or other types of impairments. In this paper, a novel approach is presented which incorporates the accessibility of images for users with various vision impairments, such as color blindness, cataract and glaucoma, in order to rerank the results of an image search engine. The accessibility of individual images is measured through the use of vision simulation filters. Multi-objective optimization techniques utilizing the image accessibility scores are used to handle users with multiple vision impairments, while the impairment profile of a specific user is used to select one from the Pareto-optimal solutions. The proposed approach has been tested with two image datasets, using both simulated and real impaired users, and the results verify its applicability. Although the proposed method has been used for vision accessibility-based reranking, it can also be extended for other types of personalization context.

Keywords Accessibility-based search engines · Accessibility reranking · Accessibility feature extraction · Multi-objective optimization

✉ Ilias Kalamaras
i.kalamaras11@imperial.ac.uk

Nikolaos Dimitriou
nikdim@iti.gr

Anastasios Drosou
drosou@iti.gr

Dimitrios Tzovaras
dimitrios.tzovaras@iti.gr

¹ Department of Electrical and Electronic Engineering,
Imperial College London, SW7 2AZ, London, UK

² Information Technologies Institute,
Centre for Research and Technology Hellas, Thessaloniki, Greece

1 Introduction

Personalization in contemporary multimedia search engines is mostly accomplished in the same manner as in search engines that retrieve web-pages: by considering meta-information [22, 32], such as the history of user queries [27] or spatio-temporal characteristics [28] in order to provide recommendations. However, the actual content of multimedia has not been widely used for personalization purposes, with few exceptions [26]. For instance, their colors or the frequencies appearing in audio data and associating them to user preferences regarding color and frequencies has not been much explored. Provided that the current paper deals with vision-related accessibility issues and without loss of generality the focus will be put upon on the analysis of images, that form a major proportion of the existing Web multimedia content. In particular, the assumption is that the content of images can be an invaluable source of information for providing enhanced personalization for people with visual impairments, via the preservation and promotion of the most accessible results. Researching in this direction is important, considering that people with vision impairments are not just a tiny minority, but their total number is estimated at about 3.8 % worldwide [39].

Personalization based on image accessibility poses two major challenges:

- In order to recommend the most accessible images for an impaired user, there is a need for an automatic procedure that quantifies how accessible the images are for people with disabilities.
- It should be taken under consideration that the search engine users may have a multitude of vision impairments simultaneously, e.g. both cataract and glaucoma.

Existing approaches for personalization have not so far considered accessibility-related information. On the other hand, methods for the automatic quantification of the accessibility of images have been developed and exhibit a potential for further usage [35]. However, so far they have targeted specific impairments and applications. There is a lack of generic approaches for the accessibility assessment of images, especially within the context of search engine personalization. Moreover, in the guidelines developed so far for the evaluation of web page accessibility, as well as in the existing methods for evaluating image accessibility, the impaired users are considered as having a single impairment [33, 35], which is not always true. In fact, a large proportion of visually impaired users have more than one coexisting vision impairments. Patients having both cataract and glaucoma are quite frequent [16]. Studies of the National Eye Institute [30] show that about 4.5 % of the overall population has some form of color blindness, while about 45 % of people around 75 years old have cataract. Since color blindness is not related to age and is independent from the development of cataract, this means that about 2 % of the population around 75 years old suffer from both cataract and color blindness. Considering a user as having multiple impairments can provide more accurate personalization than considering only a single impairment. The consideration of multiple characteristics describing the same entity has proven useful in other areas, such as machine learning [2], which motivates the use of such multimodal learning techniques in accessibility-related personalization.

In this paper, a novel approach for providing accessibility-related personalization in image search engines is proposed. Personalization is provided in the form of reranking the most relevant retrieved images, so that the most accessible ones to a specific user are promoted to the first positions. Although the proposed approach has been developed for image search engines and vision-related impairments, the introduced concepts and

procedures can be used with any type of multimedia and impairments. In order to address the aforementioned challenges, this paper makes the following contributions:

- Vision accessibility-related information is introduced for personalized search engines.
- Image accessibility is automatically computed using a novel and generic approach utilizing vision simulation filters. Although the vision filters used are based on existing filters of the literature, to the authors' knowledge, it is the first time that the accessibility of image is quantitatively evaluated through a generic procedure able to handle several types of vision impairments.
- The multitude of impairments that a user may have is handled by formulating the problem of image reranking as a multi-objective optimization problem, and selecting the optimal solution based on the specific user impairments.

The rest of the paper is organized as follows. A review of the recent literature related to the subject of accessibility-aware search engines is presented in Section 2. The ground truth dataset used for the evaluation of the proposed methods is described in Section 3. The proposed approach to extract the accessibility scores from the images is presented in Section 4, followed by the presentation of the multi-objective optimization approach, in Section 5. Section 6 contains the results of the experimental evaluation of the proposed approach. Finally, Section 7 concludes the paper.

2 Related work

The majority of work related to personalized search engines, considers the search history of the user, or the search history of other users, as context [1]. In [25], the previous query requests of the users are collected and used to provide personalized recommendations. In [6], the query history of a user is clustered in conceptual groups. The sequence of concepts submitted by many users are used as the context to provide query recommendations. Information about the previous queries and the results that were selected by the user is used also in [40] and are integrated into a ranking model. Other contextual factors, such as spatio-temporal and environmental aspects have also been used to rerank the search results, as in the works of [3] and [7].

The effect of well-known vision impairments, like cataract and macular pathology, to the visual acuity and contrast sensitivity of impaired users respectively, has been thoroughly studied and reported in the past [18, 24]. However, as demonstrated by the results of [24], with regard to elderly people with visual acuity problems, uncorrected vision is often overlooked in research about vision and refraction. There is a bias towards people wearing glasses over those not wearing them, which can affect the way that visual aids, such as street signs, are designed. The study of [14] showed that, in typed text, words are perceived clearer than individual letters, for people with visual acuity problems caused by central field loss and cataract.

With regard to web-pages, the most significant accessibility-related guidelines are included and described in Guidelines of Web Content Accessibility (WCAG) [37]. WCAG contains guidelines for making web content accessible to people with disabilities, such as providing textual or other alternatives for images and auditory content, not relying only on the colors of the web-page to provide information ensuring that the presented documents are simple and clear. After the first web accessibility guideline which was compiled by Gregg Vanderheiden in 1995, over 38 different Web Access guidelines followed from

various authors and organizations. They consist of a set of guidelines for making content accessible, primarily for disabled users, but also for all user agents, including highly limited devices, such as mobile phones.

However the attention of these guidelines is focused on web-pages and ensuring that their content can be easily read by the majority of disabled users. Concerning multimedia, only few works have addressed the problem of evaluating the accessibility of individual images, in terms of their visual content. In [35], the authors discover areas within images which are inaccessible for people with color-blindness. This is accomplished by simulating the perception of the image by the user and calculating edge differences between the original and the simulated images. In [41] and [34], the content of the images is processed in order to enhance the images, so as to be better accessible for people with color-blindness and decreased contrast sensitivity, respectively. Modifying the functionality of a web page or service, for instance of a search engine, so as to be adapted to people with disabilities, has not been given much focus yet.

3 Ground truth dataset

The purpose of the developed methods is to facilitate the use of image search engines by visually impaired users. In order to assess the effectiveness of the proposed methods, a ground truth dataset has been collected by recruiting people with vision impairments. For the collection of the ground truth data, a web-based tool has been implemented, through the use of which, impaired users are able to assess the visibility of various sets of images and to submit appropriate rankings of them. In each user session, the user is presented with a set of 10 images, in random order. The images are randomly picked from a set of 100 fashion-related images, taken from the fashion dataset of the CUBRIK project [10]. As a first task, the user is requested to put the images in order, from the one which is most easy for them to see to the one which is most difficult to see. Whether an image is easy or difficult to see is left to the user's perception, without the researcher providing any clues about items that may exist in the images, but are not seen by the user.

Once the user submits the ordering of the images, the second phase of the experiment follows. In this phase, the users are requested to check whether they see or not a number of visual characteristics appearing in the images, such as specific objects or colors. The visual characteristics for each image have been gathered from the manual annotation of the images by a number of users with full vision. The images for which the users are requested to check the visibility of the characteristics are the same as the ones which the users ordered in the previous phase of the experiment, and are presented to the users one after the other, in the same order as they had put them.

When the user completes the second phase of the experiment for all 10 images, one user session ends. Each user was requested to participate to five of the above two-phase sessions, each with another random set of 10 images. The images in each session are results of five fashion-related textual queries, namely "hat", "jeans", "shirts", "shoes" and "skirt".

In the experiments, 10 visually-impaired users have participated. The users were patients of the ophthalmological clinic of the AHEPA hospital in Thessaloniki, Greece, and of the Social Insurance Institute of Neapoli, Thessaloniki. The number of users is relatively small due to the difficulty in finding visually-impaired users for the purposes of collecting ground truth data. Of these users, 5 were women and 5 were men, while their ages ranged from 62 to 83 years old, except for one aged 34. Most patients suffered from glaucoma, in some cases along with cataract and protanopia. Two of the patients suffered only from cataract and one

only from protanomaly. Table 1 contains the characteristics of all users in detail. A user impairment profile has been created for each user, corresponding to his/her impairments and their severity.

Each patient was requested to use the ground truth collection tool for five sessions. Each session corresponded to a sample query submitted to the CUBRIK search engine. In particular, text queries, such as “jeans” and “shoes” were submitted to the search engine. The top ten resulting images were used as the images presented to the user in one session of the ground truth collection tool, in the order returned by the search engine, i.e. by decreasing relevance score. This ordering is hereby referred to as the “relevance ranking”.

For each image i presented in the user sessions, a ground truth accessibility score $a_{i,gt} \in [0, 1]$ was calculated, based on the visual characteristics that the user checked as visible. The accessibility score was calculated as the ratio of the visible characteristics over the total number of visual characteristics existing for each image:

$$a_{i,gt} = \frac{n_{\text{visible},i}}{n_{\text{total},i}} \quad (1)$$

The larger the accessibility score, the more easy was for the user to see the items appearing in the image.

4 Automatic computation of image accessibility

In this section, the proposed approach for automatically computing the accessibility of images is presented. An overview of the approach is provided, followed by detailed descriptions of its parts.

4.1 Overview

Accessibility-based reranking of the search results relies on extracting accessibility scores from the images, in analogy to the relevance scores calculated by a standard search engine.

Table 1 Characteristics of the visually impaired users participating in the evaluation

User ID	Gender	Age	Impairment(s)
1	male	66	medium cataract, medium glaucoma
2	female	83	medium cataract, medium glaucoma
3	male	66	severe glaucoma
4	female	75	mild glaucoma
5	male	70	mild cataract
6	female	70	mild cataract, severe glaucoma
7	female	69	mild cataract
8	female	72	mild glaucoma
9	male	62	mild glaucoma, protanopia
10	male	34	medium protanomaly

For each of the supported vision impairments, an accessibility score $a_m \in [0, 1]$, $m = 1 \dots M$, is extracted from an image, evaluating how accessible this image is.

The overall procedure of extracting the accessibility score a_m of an image I , for impairment m is the following. The impairment m is modeled as a filter f_m , which distorts the original image, according to vision impairment m . The filtered image, I_m , is a simulation of how an impaired user having impairment m would perceive the original image I . Next, a comparison is performed between the original and the filtered image, using a distortion measure g_m , in order to quantify the distortion imposed by the impairment filter. Images with content that is accessible to the user would undergo less distortion than images that are not accessible. For instance, let a protanopia filter be considered. The filter simulates how a person with protanopia would perceive an input image. If the image contains several red and green areas, the image would be highly distorted by the filter. On the other hand, if the image does not contain red and green areas, the image would not be distorted at all. Thus the amount of distortion imposed by the filter to the image is an indication of how accessible this image is. This amount of distortion is used as the image accessibility score a_m . Hereby, a_m is normalized in the $[0, 1]$ range, with 0 meaning that the image is not accessible and 1 meaning that it is totally accessible. A graphical overview of the procedure is depicted in Fig. 1. The novelty of this architecture, is first that it allows the extraction of a numeric score assessing the accessibility of an image for people with a specific type of impairment, and also that it is generic enough to handle any type of vision impairment, simply by using an appropriate vision filter for the impairment.

4.2 Vision filters used

Formally, a vision filter f_m is defined as a function

$$f_m : \mathcal{I} \rightarrow \mathcal{I}, \quad (2)$$

where \mathcal{I} is the space of images, which takes an image I as its input and produces another image I_m , which is the simulation of how a person having impairment m would perceive the original image. Depending on the characteristics of impairment m , the filter f_m may modify the colors of the pixels, may cause blurring of the image, etc.

In this paper, five vision impairments are considered, namely *cataract*, *glaucoma*, *protanopia*, *deutanopia* and *tritanopia*. The filters used for the five considered impairments are described in the following sub-sections. Most of the filters are based on the ones used in the vision simulation system of [17]. The filters of [17] have been developed and evaluated within the context of the European project ACCESSIBLE [12], which focused

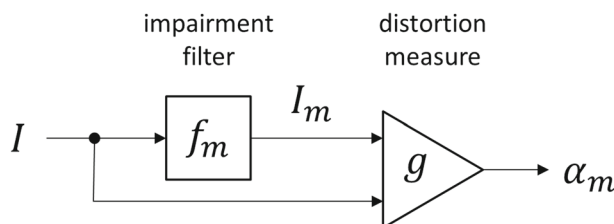


Fig. 1 Overview of accessibility score extraction for vision impairment m

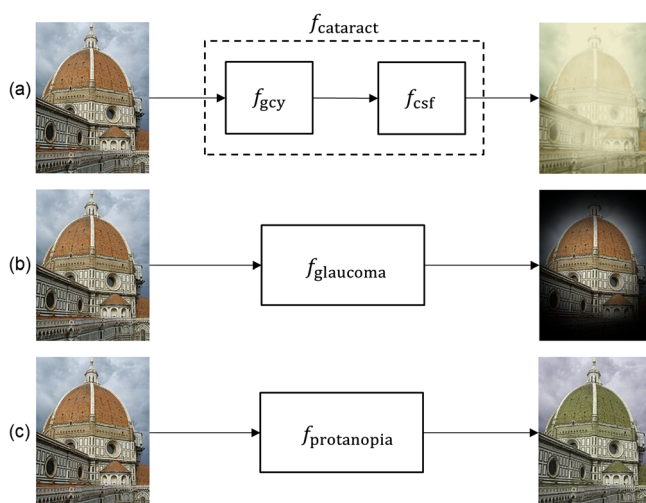


Fig. 2 Application of the filters for the considered impairments on an example image. **a** Application of the cataract filter. The cataract filter consists of the glare, clouding and yellowing filter of [17], followed by the contrast sensitivity and visual acuity filter based on the CSF function. **b** Application of the glaucoma filter. **c** Application of the protanopia filter

on the simulation of various types of impairments, for the purpose of designing accessible applications and interfaces. Figure 2 collectively illustrates the effect of the filters to an input image.

4.2.1 Cataract

Cataract is the deterioration of vision due to the clouding of the lens. As stated by the National Eye Institute [30], the effect of cataract can be analyzed in a series of simpler effects, including the following:

- decreased sensitivity in low contrast changes
- decreased visual acuity (blurriness)
- increased sensitivity to glare sources (bright areas in the image)
- perception of bright clouds (like “cataracts”) in the visual field
- yellowing of the image

The filter simulating cataract, f_{cataract} , is herein split into two sub-filters, f_{gcy} and f_{csf} , each responsible for a different group of cataract effects. The first sub-filter, f_{gcy} , simulates the glare sensitivity, clouding and yellowing symptoms, and makes use of the filters of [17]. In [17], glare sensitivity is simulated through manipulations of the bright areas in the image and their exaggeration, so that they affect large part of the visual field. The clouding effect is approached using randomized semi-opaque masks, simulating *scotomata*, i.e. areas resembling clouds or cataracts, covering the whole visual field. Finally, the yellowing of the image is simulated by modifying the hue and saturation of the image in the HSV color space. The intensity of all these effects can be varied within a range of no effect to full effect. In particular, glare sensitivity is controlled by varying the size of the area affected by bright spots, clouding is controlled by varying the transparency and size of the scotomata and yellowing is controlled by the amount of hue and saturation distortion.

The second sub-filter of cataract, f_{csf} , simulates the contrast sensitivity and visual acuity effects. In this case, a unified approach, different from [17], is followed, so as to produce more realistic results. The approach is based on the Contrast Sensitivity Function (CSF) of the human eye. Contrast is generally defined as the luminance difference between two points in an image, normalized by the average image luminance. There is a threshold in the contrast values that the human eye can see. Luminance differences smaller than this threshold cannot be perceived. However, this threshold varies with the spatial frequency of the luminance source. This variation of the contrast threshold, or usually its inverse, the *contrast sensitivity*, is described by the Contrast Sensitivity Function (CSF).

The CSF can be formulated as a band-pass filter [19, 20, 34] with a peak at middle frequencies. Adopting the formulation of [34], the contrast sensitivity function for an impaired user, considering middle to high frequencies, can be modeled by the following exponential function:

$$\text{CSF}(u) = (1 - L_c)e^{-0.166u/(1-L_d)}, \quad (3)$$

where u is the magnitude of the spatial frequency in degrees per visual angle and $L_c \in [0, 1]$ and $L_d \in [0, 1]$ are parameters modeling the impairment of the user. The L_c parameter models the contrast sensitivity of the user. Larger values of L_c denote a lower contrast sensitivity, meaning that the user cannot distinguish small differences in intensity. This is modeled by scaling down the magnitude of the CSF in (3), as L_c grows. The L_d parameter models the visual acuity of the user. Larger values of L_d denote a poorer visual acuity, meaning that the user cannot see details of high spatial frequency. This is modeled by shrinking the CSF function of (3) along the frequency axis, as L_d grows. This type of analysis, with the use of CSFs, incorporates both contrast sensitivity and visual acuity, and is thus closer to the actual user perception, than separately reducing the image contrast and blurring it, as in [17]. The application of the contrast sensitivity and visual acuity filter can be performed by multiplying the CSF function of (3) to the magnitude of the frequency spectrum of the image.

The complete cataract filter consists of the application of the glare, clouding and yellowing sub-filter, f_{gcy} , followed by the contrast sensitivity and visual acuity filter f_{csf} . The application of the cataract filter to a sample input image is as illustrated in Fig. 2a.

4.2.2 Glaucoma

Glaucoma is the damage of the optic nerve, usually related to increased fluid pressure in the eyes. The effect of glaucoma on the human vision is the perception of a dark area around the center of vision. The size and intensity of the dark area depends on the severity of the impairment.

For the glaucoma filter, f_{glaucoma} , the implementation of [17] has been used hereby. In this implementation, glaucoma is simulated using a semi-opaque circular black mask applied at the periphery of the visual field. Different from the scotomata masks used in the cataract filter, the mask used for glaucoma covers the periphery of the visual field, leaving the center of vision unaffected. The severity of the impairment can be varied by controlling the size and the transparency of the peripheral area covered by the mask. These range from zero size and full transparency, i.e. normal vision, to almost the size of the visual field and full opacity, i.e. almost total blindness. Figure 2b illustrates the effect of applying the glaucoma filter on an example image.

4.2.3 Protanopia, deuteranopia, tritanopia

Color blindness is the inability to see adequately or at all a region of the visible light spectrum. It is caused by faults or total absence of one or more retinal photoreceptor cones, which are responsible for perceiving color and transmitting the information to the optical nerve. In case of a single deficient cone, and depending on the color for which the deficient cone is responsible, color blindness is split into three types: protanopia, related to the red cone, deuteranopia, related to the green cone, and tritanopia, related to the blue cone. People with protanopia and deuteranopia, which are the most usual types of color blindness, have difficulty in distinguishing between red and green colors, while people with tritanopia have difficulty in perceiving blue colors. If the respective cone is not totally absent, but it rather faulty, causing a mild impairment, the three above types of impairments are called, respectively, protanomaly, deuteranomaly and tritanomaly.

For the color blindness filter, $f_{\text{protanopia}}$, $f_{\text{deuteranopia}}$ and $f_{\text{tritanopia}}$, the implementations of [17] are again used. In [17], filters for all types of color-blindness are defined and used. The same filters are also used in this paper. The filters are implementations of the color-blindness simulators described in [5]. In particular, the input image is transformed from the RGB color space to the LMS (Long, Medium, Short) color space [13], which represents colors with respect to the response of the three types of cones in the retinal photoreceptor to light of long, medium and short wavelength. This transformation is due to the fact that the relationship between the components of the RGB color model and their perception by the three types of cones is not linear, but there is a rather large overlap between the responses of the red and green cones, while the blue one is separated. Protanopia is simulated by modifying the value corresponding to the long wavelength cone, i.e. the one related to the red color, so that the responsiveness to the red color is eliminated. Deuteranopia and tritanopia are simulated similarly, for their corresponding wavelengths. The final image is acquired by transforming back to the RGB color space. Various degrees of protanomaly, deuteranomaly and tritanomaly are simulated by considering the final image as a weighted average of the original image and the one corresponding to complete protanopia, deuteranopia and tritanopia, respectively, and controlling the weight. Thus, e.g., a weight of 0 for protanopia would correspond to normal vision, while a weight of 1 for protanopia to complete protanopia. In Fig. 2c, the application of the protanopia filter on a sample image is depicted, as an example of a color blindness filter.

4.3 Distortion measures used

After the calculation of the filtered image, for either of the five impairments, a distortion function is used to compare the original and the filtered image. The distortion function measures the information loss caused by the filtering procedure.

A distortion function g_m is defined as

$$g_m : \mathcal{I} \times \mathcal{I} \rightarrow [0, 1], \quad (4)$$

i.e. it takes two images as its input and computes a value in the range $[0, 1]$ as its output, measuring the distortion of the second input image, compared to the first. Various functions can be used as distortion functions, for instance information loss, difference of color histograms or difference of detected edges. Different distortion functions may be suitable for different impairments.

The distortion function used for all impairments is the sum of three distortion measures, measuring differences in:

- the luminance histogram,
- the detected edges and
- the pixel-by-pixel color values of the images.

In the literature, differences in luminance, edges/gradients and color have commonly been used for the evaluation of quality degradation of an image with respect to a reference image [36, 42]. These types of degradations are based on characteristics of the human visual system (HVS) perception, since, while an image may contain redundant visual information, the HVS uses only low-level features, such as edges, in order to perceive it [42]. Hereby, the differences in luminance between the original and the filtered images are captured by the differences in the luminance histograms, differences in gradients are captured by the differences in the detected edges and differences in color are captured by the differences in the pixel-by-pixel color values. Other implementations of distortion measures to capture these three types of distortions could of course be used as well. However, the focus of this paper is mostly on demonstrating that such an approach and architecture can prove valuable, rather than searching for the most fine-tuned distortion measures.

4.3.1 Difference of histograms

For the difference of histograms, the images are first transformed to the *Luv* color space and then only the luminance (*L*) channel is used. The luminance histograms \mathbf{h}_I and \mathbf{h}_{I_m} of the original image *I* and the filtered image *I_m*, respectively, are then constructed, quantizing the luminance values in *b* bins. Let $h_{I,k} \in [0, 1]$, $k = 1 \dots b$, be the value of the \mathbf{h}_I histogram in the *k*th bin, and similarly for \mathbf{h}_{I_m} . The histogram values are normalized so that $\sum_{k=1}^b h_{I,k} = 1$, and similarly for \mathbf{h}_{I_m} . Then, the histogram difference distortion function, g_h , is defined as the Euclidean distance between \mathbf{h}_I and \mathbf{h}_{I_m} :

$$g_h(I, I_m) = \sqrt{\sum_{k=1}^b (h_{I,k} - h_{I_m,k})^2} \quad (5)$$

For the implementation of this function, the histograms have been considered to consist of 64 bins, i.e. $b = 64$.

4.3.2 Difference of edges

For the difference in the edges detected in the original image *I* and the filtered one *I_m*, the images are again transformed to the *Luv* color space and only the luminance channel is considered. Then, the Sobel edge detection operator is applied to both of them, producing images *I_s* and *I_{m,s}*, respectively. Images *I_s* and *I_{m,s}* are gray-scale images, where the intensity of each pixel, normalized to the [0, 1], indicates if there are edges at this position in the original and filtered images. The closer the intensity is to 1, the sharper an edge exists in the respective initial image.

The amount of edges in the original and filtered images can be approximated by summing the intensity values of the images produced after edge detection. Let I_s^{ij} be the intensity of the pixel in the *i*th row and the *j*th column of image *I_s*, and similarly for *I_{m,s}*. Let also e_I and e_{I_m} be the amount of edges in images *I* and *I_m*, respectively. Then

$$e_I = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W I_s^{ij}, \quad (6)$$

$$e_{I_m} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W I_{m,s}^{ij}, \quad (7)$$

where H and W is the height and the width of the images, respectively.

The edge-related distortion function, g_e , can then be defined as the normalized difference between s_I and s_{I_m} :

$$g_e(I, I_m) = \frac{e_I - e_{I_m}}{e_I}. \quad (8)$$

4.3.3 Pixel-by-pixel difference

The third distortion function, g_p , measures the pixel-by-pixel difference in color between the original and the filtered images. Similar to the previous functions, the distortion function g_p takes two images I and I_m as its input. Let $I^{ij,R}$, $I^{ij,G}$ and $I^{ij,B}$ be the R , G and B color components of the pixel in the i th row and j th column of image I , and similarly for I_m . Let also $d_{ij}^2(I, I_m)$ be the squared color difference between the pixels in the (i, j) position in the I and I_m images:

$$d_{ij}^2(I, I_m) = (I^{ij,R} - I_m^{ij,R})^2 + (I^{ij,G} - I_m^{ij,G})^2 + (I^{ij,B} - I_m^{ij,B})^2 \quad (9)$$

The distortion function is defined as follows:

$$g_p(I, I_m) = \frac{1}{HW} \frac{1}{B} \sum_{i=1}^H \sum_{j=1}^W c_{t,ij}(I, I_m) d_{ij}^2(I, I_m), \quad (10)$$

where H and W are the height and the width of the image, respectively, in pixels,

$$B = \sum_{i=1}^H \sum_{j=1}^W c_{t,ij}(I, I_m) \quad (11)$$

is a normalization constant, and

$$c_{t,ij}(I, I_m) = \begin{cases} 1, & \text{if } d_{ij}^2(I, I_m) > t \\ 0, & \text{otherwise} \end{cases}, \quad (12)$$

is a parameter introduced to keep only the differences that are larger than a threshold value t . This thresholding has been introduced in order to ignore small differences in color that are due to noise introduced by the filter. The specific value of the threshold is determined so that it corresponds to a small percentage, hereby 1 %, of the range of the difference values in the image.

4.3.4 Total distortion

Finally, the total distortion function, g is defined as the weighted sum of g_h , g_e and g_p :

$$g(I, I_m) = w_h g_h(I, I_m) + w_e g_e(I, I_m) + w_p g_p(I, I_m), \quad (13)$$

$$w_h, w_e, w_p \in [0, 1], \quad w_h + w_e + w_p = 1.$$

A linear combination of the three distortion measures has been used due to its simplicity and capability of assigning significance weights to the individual distortion measures. Different kinds of combinations can also be considered, which can be a direction for future research. The weights w_h , w_e and w_p , above, determine the trade-off between the histogram, edge and pixel-by-pixel functions for the calculation of the final distortion function. For the implementation used in this paper, equal weights have been assigned to all distortion functions,

i.e. $w_h = w_e = w_p = 0.33$, in order to consider them with equal significance. Experimentation with weights around these values have shown that the results are not significantly affected, thus equal weights have been selected.

Overall, introducing in the notation for images I and I_m the subscript $i \in \{1 \dots N\}$, denoting one of the N images returned by the search engine, the accessibility score for the image with index i and for impairment m is calculated as follows:

$$a_{i,m} = g(I_i, I_{i,m}) = g(I_i, f_m(I_i)), \quad (14)$$

$$m \in \{\text{cataract, glaucoma, protanopia, deuteranopia, tritanopia}\}.$$

It should be noted that the accessibility score for any individual impairment is computed using the total distortion function, i.e. a combination of all three distortion measures. There is no correspondence between a distortion measure and an impairment. Each of the three distortion measures captures a different type of distortion that may be imposed due to any of the five impairments. For instance, filtering an image showing a red object in a green background with the protanopia filter would make the object difficult to distinguish from the background. This would cause an amount of distortion in all three types of distortion measures: in the difference of luminance histograms, since the luminance of the object in the filtered image is closer to the luminance of the background, in the difference of edges, since the edges of the object are not clear in the filtered image, and in the pixel-by-pixel color difference, since the colors of the filtered image are different than the original one.

4.4 Automatic accessibility scores and ground truth validation

As a first validation of the above procedure for extracting the accessibility scores from the images, the automatically extracted scores have been compared to the ground truth accessibility scores described in Section 3. The experimental setting is as follows. Each of the recruited users presented in Table 1 were considered in turn. Each user has a specific impairment or combination of impairment in various amounts. The vision of the user was simulated as a combination of all the filters described in Section 4.2, where each filter was tuned in order to agree with the amount of disability of the user for the corresponding impairment. Thus, an input image passing through all the appropriately tuned filters would produce a distorted image that simulates how the user, with the specific combination of impairments, sees the image. Then, the distortion between the original and the final image is calculated, in order to extract an accessibility score for the image, as described in Section 4.3.

In this manner, accessibility scores for all the images in all sessions described in Section 3 have been computed, each time tuning the filters according to the impairment amounts of the user participating in the session. Then, the correlation between the automatically extracted scores and the ground truth accessibility scores, also described in Section 3, has been calculated, using the Pearson correlation coefficient. The correlation value was 0.2027, suggesting a positive correlation between the automatically extracted scores and the ground truth scores, which indicates that the proposed method manages to agree with the user perception.

However, the approach of passing an image through all filters and computing the distortion is computationally heavy and would cause an online search engine system to perform slowly, since the accessibility scores of all returned images for a specific user, with specific impairment amounts, need to be computed on the fly. A different, more efficient, approach would be to pre-compute the accessibility scores of the images for the separate impairments, produce an optimal ranking for each impairment separately, and then considering the exact

user impairment amounts, in order to merge the rankings, e.g. using a weighted sum, putting more focus on the ranking corresponding to the impairment for which the user has the largest disability. Computing the optimal ranking for a specific user can be considered as an optimization problem, where the objective to be optimized is the appropriateness of a ranking for the specific user. Using multiple rankings of the results and merging them is equivalent to considering combinations among multiple optimization objectives. However, combining the objectives in a weighted sum fashion and optimizing the combined objective has the disadvantage that there may be appropriate solutions, i.e. rankings, that cannot be discovered, even if all weight combinations are considered for the objectives, leading to the need for more elaborate methods [23]. Hereby, this problem is addressed using multi-objective optimization techniques, as presented in the following.

5 Incorporation of multiple vision impairments using multi-objective optimization

In this section, the multi-objective approach used in order to handle the simultaneous existence of multiple vision impairments in a user is described.

5.1 Overview

Using the accessibility score extraction procedure described in Section 4, accessibility scores for each image of the search results can be calculated for the supported vision impairments. A first issue arising here is the amount of distortion imposed by the filters. The nature of the vision impairments is that a user does not either have or not have an impairment, but rather has an impairment at an amount within a continuous range of degrees. The vision filters used are able to simulate this by varying their intensity level. Thus, the amount of distortion imposed in the input image is affected not only by the characteristics of the image itself, but also by the intensity of the filter. In order to compensate for this fact, the intensity of the imposed filters is fixed to the average intensity of each filter's intensity range, as if they corresponded to an impaired user having an average amount of disability for each of the supported impairments. This consideration does not actually affect the relative accessibility between two images, with respect to a single impairment, since impaired persons would perceive an accessible image better than an inaccessible one, regardless of the amount of disability they have. It has an impact, however, when considering the combination of multiple impairments. This is the reason for selecting the average amount of intensity for the filters of all impairments, in order for the set of Pareto-optimal rankings computed, described below in this section, to be balanced across the impairments. The actual amount of disabilities of the impaired users in each of the supported impairments is taken into account later, for the final selection of a single Pareto-optimal solution, as described in Section 5.3.

If a single impairment was considered, a simple ordering of the images in order of descending accessibility scores would be sufficient for an accessibility-based reranking. However, in this paper, the users are considered to have more than one impairments, so that ordering the images with respect to one impairment conflicts with their ordering with respect to another.

Such problems of conflicting objectives can be handled using multi-objective optimization techniques [9, 11]. Multi-objective optimization deals with trying to simultaneously optimize a set of conflicting objectives and results in a set of optimal trade-offs among the multiple objectives, called the Pareto-optimal solutions. Herein, an approach similar to

the multi-objective visualization method of [21] is adopted, with the conflicting objectives being the orderings of the results according to either of the considered impairments.

Formally, the multi-objective ranking problem is hereby stated as follows. Let \mathcal{F} be the set of all possible rankings of the results. Let also

$$\mathbf{J} = \{J_1, \dots, J_M\}, \quad J_m : \mathcal{F} \rightarrow \mathbb{R}, \quad m = \{1 \dots M\}, \quad (15)$$

be a set of objective functions, each evaluating, with a numerical score, the appropriateness of a particular ranking $p \in \mathcal{F}$ with respect to a specific criterion, which is hereby related to the different vision impairments. The goal of multi-objective optimization is to simultaneously minimize all objectives:

$$\min_{p \in \mathcal{F}} \mathbf{J}(p) \quad (16)$$

Instead of resulting in a single solution, as with single-objective optimization problems, this multi-objective optimization results in the set of Pareto-optimal solutions $\mathcal{P} \subseteq \mathcal{F}$. Multi-objective optimization is based on the notion of *dominance* among the possible solutions. A particular solution dominates another one, if it has a smaller value in at least one of the objectives, without having a larger value to any other objective. In other words, it is at least better than the other in one objective, without sacrificing any other objective. The set of Pareto-optimal solutions is the set of solutions that dominate all other ones without dominating each other. One of the Pareto-optimal rankings $p \in \mathcal{P}$ is finally selected, based on the specific impairment profile of the user.

5.2 Objective functions used

The *discounted cumulative gain* (DCG) is used for the definition of the objective functions. The DCG is commonly used for the evaluation of the effectiveness of the rankings produced by search engines [29]. Considering that a set of N results is ranked, so that result i has a rank r_i , with 1 being the top-most rank, the DCG is calculated as follows:

$$\text{DCG} = \sum_{i=1}^N s_i, \quad s_i = \begin{cases} \text{rel}_i, & r_i = 1 \\ \frac{\text{rel}_i}{\log_2 r_i}, & r_i > 1 \end{cases} \quad (17)$$

where rel_i is the relevance score of image i . The larger the DCG, the more promoted (i.e. placed in higher positions) are results that are relevant to the query.

The DCG is used hereby as an objective function for evaluating the rankings of the results, with the modification that the relevance score is replaced with the accessibility scores for the various supported impairments. If $\mathbf{r} = (r_1, r_2, \dots, r_N)$ is a ranking of the N results returned by the search engine, then the objective function corresponding to impairment m is the following:

$$J_m(\mathbf{r}) = 1 - \frac{1}{N} \sum_{i=1}^N t_i, \quad t_i = \begin{cases} a_{i,m}, & r_i = 1 \\ \frac{a_{i,m}}{\log_2 r_i}, & r_i > 1 \end{cases} \quad (18)$$

where $a_{i,m} = g(I_i, f_m(I_i))$, is the specific accessibility score of image i , for impairment m . The DCG has been normalized by the number of results N and subtracted from 1, in order for the optimal ranking to be calculated by minimizing the objective function, instead of maximizing them. Hereby, five objective functions of the above form are defined, one for each of the considered impairments, i.e. $m \in \{\text{cataract, glaucoma, protanopia, deuteranopia, tritanopia}\}$. Note that the DCG metric is also used later, in Section 6.2, for evaluating the effectiveness of the proposed method, based on

ground truth. In that case, the extracted accessibility scores $a_{i,m}$ of the images are replaced by the ground truth accessibility scores $a_{i,gt}$, presented in Section 3.

Using the DCG objective functions, all possible rankings of the results can be evaluated for each of the five impairments and the Pareto-optimal ones can be calculated using multi-objective optimization techniques. Hereby, the SPEA2 genetic algorithm [43] is used for the calculation of the Pareto front. Genetic algorithms are commonly used for solving multi-objective optimization problems, since the fact that they maintain a population of solutions, instead of a single one, makes them more appropriate for computing the Pareto-optimal set. Dominance relations are commonly used as the fitness functions. SPEA2 (Strength Pareto Evolutionary Algorithm 2), instead of considering whether a solution dominates other solutions, measures how many members of the population a solution dominates or is dominated by, an amount named as *strength*. The fitness function used is based on this strength value, rather than on mere dominance. SPEA2 maintains, apart from the regular solution population, an archive of the most dominant solutions that have appeared throughout all the iterations up to the current one, ensuring that no good solutions are missed due to random effects. The size of the archive is kept constant, using clustering-based truncation operations, in order to achieve a uniform density of the archived solutions. Members from both the regular population and the archive are used for the evaluation of the fitness values and for the recombination and mutation operations.

After the calculation of the Pareto-optimal rankings, one of them needs to be selected, in order to be presented to the user, as described in the following.

5.3 Selection of a single solution based on the user impairment profile

All the aforementioned impairments can have variable amounts of severity. In order to model the amount of disability of a user in a specific impairment m , a decimal value $x_m \in [0, 1]$ is used. In this paper, this value will be referred to as the impairment amount, for impairment m . The impairment amount takes values in the range from 0 to 1, with 0 meaning that the user does not have impairment m at all, and 1 meaning that the user has the impairment in the largest possible amount. A user with having an average severity for impairment m is considered as having an impairment amount of $x_m = 0.5$.

An impaired user may have a multitude of impairments simultaneously, possibly at different impairment amounts each. For instance, the user may have both cataract and glaucoma, or more impairments simultaneously. Thus, instead of a single value characterizing the impairment of a user, a vector of values is used:

$$\mathbf{x} = (x_1, x_2, \dots, x_M), \quad x_m \in [0, 1], \quad m = 1 \dots M, \quad (19)$$

where M is the number of supported impairments. In this paper, $M = 5$, since five impairments are considered. This vector of impairment amounts for each of the considered impairments is hereby referred to as the *user impairment profile* and fully describes the impairments of a specific user.

For the selection of one of the Pareto-optimal solutions, the specific impairment profile of the user is used. Using the values of the impairment profile as coordinates, the user impairment profile \mathbf{x} can be positioned in the same space as the Pareto-optimal rankings. This allows the selection of the final ranking for this user as the Pareto-optimal ranking that is closer to the user impairment profile. The Chebyshev distance between the profile and the points of the Pareto front is used for this purpose. The Chebyshev distance is commonly used in achievement function-based multi-objective optimization methods, which compute

the Pareto front solutions by comparing to a reference solution [31]. The Chebyshev distance between two vectors $\mathbf{x} = (x_1, x_2, \dots, x_M)$ and $\mathbf{y} = (y_1, y_2, \dots, y_M)$ is defined as:

$$d_{CH}(\mathbf{x}, \mathbf{y}) = \max_m (|x_m - y_m|). \quad (20)$$

Let $\mathbf{J}(p) = (J_{\text{cataract}}(p), J_{\text{glaucoma}}(p), J_{\text{protanopia}}(p), J_{\text{deuteranopia}}(p), J_{\text{tritanopia}}(p))$. The selected ranking, p_{opt} is calculated as follows:

$$p_{\text{opt}} = \arg \min_{p \in \mathcal{P}} d_{CH}(\mathbf{x}, \mathbf{J}(p)). \quad (21)$$

After a specific ranking is selected, the images can finally be ordered according to it and presented to the user.

6 Experimental evaluation

The experimental evaluation of the accessibility-based multi-objective reranking method has been performed by utilizing both simulated and real impaired users, using two different image datasets. Experimentation with simulated users allows for a more controlled environment as well as a qualitative assessment of the method, while comparing with the perception of real users allows for an assessment of the method in real-world scenarios and a quantification of the results.

6.1 Evaluation with simulated users

As a first experiment for the evaluation of the accessibility-based reranking method, a dataset of 14820 images of Italian monuments, collected from Flickr, as part of the CUBRIK project [10], was used. Each image is associated with textual information, in the form of a title and tags, which can be used by a text-based search engine for image retrieval. A Solr-based search engine was used for image search and retrieval. The 10 top results are considered for accessibility-based reranking. For this use case, cataract and protanopia have been used as the supported impairments.

In Fig. 3, a set of rerankings of the results of an example query are presented, for three artificial users: one having cataract (b), one having protanopia (c) and one having both cataract and protanopia (d), with $\mathbf{x}_{\text{cataract}} = 0.5$ and $\mathbf{x}_{\text{protanopia}} = 0.3$. As a query, the word “palace” is submitted.

The first row (a) shows the the original ranking of the results, as returned by the text-based search engine, ordering the results according to their relevance to the query. Below each image, the accessibility scores extracted from the images for cataract and protanopia, using the methodology of Section 4, are presented.

For each user, Fig. 3 depicts various rankings of the results, with the ordering going from the left (top results) to the right. The first row in each user is the accessibility-based ranking of the results, computed using the multi-objective reranking method, with the accessibility scores of the images and the values of the user impairment profile. For demonstration purposes, the second row contains a simulation of how the user perceives the results of the first row. Moving from the left-most image to the right, the results are harder to perceive. This means that the ranking indeed promotes results which are easier to see for a vision-impaired user.

The advantage of the proposed procedure compared to simply ordering the results with respect to the accessibility scores of a single impairment is that the final ranking considers



Fig. 3 Example application of the multi-objective accessibility-based reranking. The query submitted is “palace”. The result images are sorted from left to right. **a** The original relevance-based ranking of the results. Below each image, its accessibility scores for cataract and protanopia are presented. In the following sub-figures, the reranking of the results for three users is depicted: **b** a user with cataract, **c** a user with protanopia and **d** a user with both cataract and protanopia. For each user, the first row is the reranking of the results, based on the accessibility scores of the images and the user impairment profile. The images of the second row are simulations of how the user would perceive the results of the first row. Images that are at the top (left) positions of the list are easier to perceive than images at the bottom (right)

all the impairments of the user, so that the first results are the most accessible with respect to e.g. both cataract and protanopia. Comparing rows (b) and (d), the ranking of row (b), which only considers cataract, promotes the most accessible images for cataract. However, if this ranking was presented to the user having both cataract and a slight protanomaly, the first image would not be much accessible, since it contains red colors. This is reflected in the protanopia score for this image, which is rather low (0.453). Instead, the proposed approach presents the ranking of row (d), in which the first image is more accessible with regard to protanopia, while still being accessible for cataract users. Moreover, by considering the specific impairment amounts of the user for each impairment, in order to guide the selection of the final solution, the relative importance of the impairments is considered, in order for the ranking to focus more on the most severe impairment.

A second round of experiments has also been conducted with simulated users, in order to evaluate the effectiveness of the proposed approach compared to existing methods of merging multiple rankings. Let N images be considered, along with their

accessibility scores $a_{i,m}$, $i \in \{1 \dots N\}$, with respect to impairment m . A ranking $\mathbf{r}_m = (r_{1,m}, r_{2,m}, \dots, r_{N,m})$, $r_{i,m} \in \{1 \dots N\}$, of the images with respect only to impairment m can be constructed, by ordering the images by their accessibility scores, in descending order, so that $r_{i,m} = 1$ means that image i is the one with the highest accessibility score for impairment m .

Such single-impairment rankings have been computed for the supported impairments, leading to a set of different rankings for the same set of images. Each image i has multiple rankings $\{r_{i,\text{cat}}, r_{i,\text{glau}}, r_{i,\text{prot}}, r_{i,\text{deut}}, r_{i,\text{trit}}\}$. In [4, 15], various methods for combining multiple rankings of a set of search results are presented. The purpose of each method is to compute a combined ranking for an image, based on its multiple individual rankings. Hereby, the following three combinations are considered:

– **max combination**

$$r_{i,\text{max}} = \max_m \{r_{i,m}\} \quad (22)$$

– **sum combination**

$$r_{i,\text{sum}} = \sum_m r_{i,m} \quad (23)$$

– **product combination**

$$r_{i,\text{product}} = \prod_m r_{i,m} \quad (24)$$

The experimental setting was as follows. A set of 20 images have been randomly picked from the fashion dataset of the CUBRIK project [10], and their accessibility scores with respect to the various impairments were computed. Then, multiple impaired users were simulated, each with an impairment profile $\mathbf{x} = (x_1, x_2, \dots, x_M)$, by considering all possible combinations of the impairment amounts, with a step of 0.1 in the amount of each impairment. For each simulated user, the proposed multi-objective ranking was calculated, along with the max, sum and product combinations of the single-impairment rankings. For each ranking, the values of the objective functions J_m for each impairment were calculated, according to (18). Then, in order to compare between the multi-objective and the combination rankings, the following measure has been used:

$$J_{\text{comp}} = \sum_{m=1}^M x_m J_m. \quad (25)$$

In other words, the value used to compare two rankings is the sum of the objectives of the various impairments, weighted by the impairment amounts of the users, so that more focus is given to those impairments where the user has the largest amount. Comparing two rankings, the one with the lowest J_{comp} measure is closer to the needs of the corresponding simulated user.

The results of the experiment are depicted in Fig. 4. Each point in the horizontal axis represents a different simulated user, with specific impairment amounts. The vertical axis is the J_{comp} value for the various rankings. The values have been ordered in increasing J_{comp} value of the multi-objective ranking. Figure 4a shows the results for all the simulated users, while Fig. 4b illustrates a zoomed region, for clarity, since the original diagram covers a wide range for the J_{comp} values. It can be observed that the multi-objective rankings consistently outperform the max ranking, having a smaller J_{comp} value for almost all simulated users, specifically for 99.92 % of them. The values of the sum and product rankings

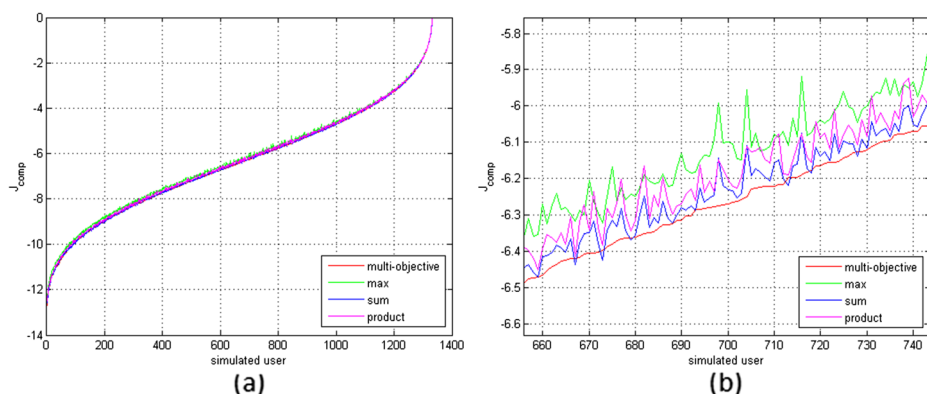


Fig. 4 Comparison of the J_{comp} values of the multi-objective and the max, sum and product rankings. The horizontal axis corresponds to different simulated users. The values have been ordered in increasing J_{comp} value of the multi-objective ranking. **a** The values for all simulated users. **b** The same plot, zoomed in a small region, for clarity

are closer to the multi-objective ranking, however, the multi-objective ranking outperforms them in 90.16 % and 99.55 % of the simulated users, respectively. These results indicate that the multi-objective rankings follow the user perception more precisely than existing methods for ranking combination.

As a qualitative example, Fig. 5 presents the five top-ranked results for a user with 1.0 protanopia, 0.1 cataract and 0.2 glaucoma. Figure 5a shows the original relevance-based



Fig. 5 The five top-ranked images for a user with 1.0 protanopia, 0.1 cataract and 0.2 glaucoma. **a** The original, relevance-based ranking. **b** The max ranking. **c** The sum ranking. **d** The product ranking. **e** The multi-objective ranking

ranking, while Fig. 5b–c depict the max, sum, product and multi-objective rankings. It can be observed that the multi-objective ranking has put black-and-white images first, which are those that can most accurately be seen by the specific simulated user, who has mostly protanopia. This ranking is closer to the user perception than the rankings produced by the max, sum and product methods.

6.2 Evaluation with real users

Experiments have also been conducted using the ground truth collected from the user sessions presented in Section 3. The images of a session were provided as input to the accessibility-based multi-objective ranking method, and a set of Pareto-optimal rankings was calculated, considering all the three impairments (cataract, glaucoma and protanopia) of the users. In order to select one of the rankings, a user profile corresponding to the impairments of each user and their amounts was created and used with the Chebyshev distance-based selection strategy. The resulting ranking, hereby denoted as “multi-objective ranking”, was compared to the relevance-based ranking, in order to assess which is closer to the user perception. The user perception has been encoded in the ground truth accessibility scores of the images, so these scores are used in the comparison, as described below.

In order to compare the two rankings, the Discounted Cumulative Gain (DCG) metric has been used. As already mentioned, given a ranked list of N search engine results, each with a rank r_i and with a relevance score rel_i , DCG is a means to quantify whether the top results in the list have high relevance scores or not. The traditional DCG is calculated as in (17). The larger the DCG, the more accurate the ranking of the results, according to the relevance scores. Hereby, instead of the relevance scores rel_i , the ground truth accessibility scores have been used:

$$DCG = \sum_{i=1}^N t'_i, \quad t'_i = \begin{cases} a_{i,gt}, & r_i = 1 \\ \frac{a_{i,gt}}{\log_2 r_i}, & r_i > 1 \end{cases} \quad (26)$$

It is important to note that the DCG metric is hereby used only for evaluation purposes. In Section 5.2, the DCG metric was used for the definition of the objective functions used

Table 2 Relevance-based and multi-objective rankings for an example session

Position	Relevance		Multi-objective	
	Image ID	Score	Image ID	Score
1	97	0.44	152	0.75
2	130	0.67	101	0.60
3	101	0.60	130	0.67
4	121	0.55	121	0.55
5	148	1.00	148	1.00
6	99	0.50	146	1.00
7	140	1.00	142	1.00
8	142	1.00	97	0.44
9	152	0.75	140	1.00
10	146	1.00	99	0.50
DCG	3.13		3.66	

in the multi-objective optimization. It used the automatically extracted accessibility scores of the images to compute the Pareto-optimal rankings and the final ranking. Hereby it is used again, this time using the ground truth accessibility scores of the images, in order to evaluate the final ranking computed automatically. The same metric is also computed for the relevance-based ranking for comparison.

For each session, the DCG evaluation metric has been used to compare the relevance ranking to the automatic ranking. The results for an example session are presented in Table 2. The query used for this example is “shirts” and they correspond to a user with a large amount of glaucoma and a small amount of cataract. The relevance and automatic rankings are presented. The image ID columns contain the unique IDs identifying the images, and are used in order to demonstrate how the positions of the images in the list differ between the relevance-based and the accessibility-based rankings. The score columns contain the ground truth accessibility scores of the images. The desired promotion of results

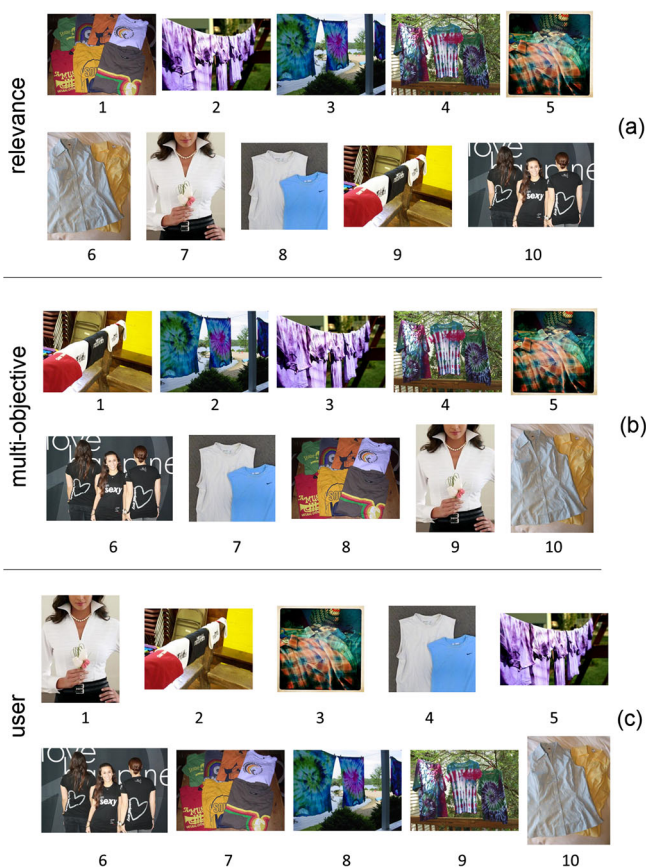


Fig. 6 Illustration of the rankings of Table 2. **a** The results of the search engine, using “shirts” as the query, ranked according to their relevance to the query. **b** The images are automatically ranked using the multi-objective method, for a user with glaucoma and cataract. Images with vivid colors and sharp edges, such as image 1, which are easier for the user to see, have been promoted. **c** The images are ranked according to their ground truth accessibility scores for this user. Images with vivid colors and sharp edges have been put at the top positions

with high accessibility scores is apparent both from the result positions in Table 2 and from the comparison of the DCG value, which is larger for the automatic ranking.

An illustration of the above rankings is presented in Fig. 6. The use of the accessibility filtering promotes images which are easier for a person with glaucoma and cataract to see, i.e. images having higher contrast and more vivid colors. For comparison, the bottom part of Fig. 6 contains the ground truth ranking of the images, according to the user annotations.

An average DCG value of 3.62 has been measured for the relevance ranking, averaged over all sessions. Using the automatic ranking, calculated by the accessibility filtering pipeline, an average DCG value of 4.09 has been measured, being by 0.47, or 12.98 %, larger than the DCG of the relevance-based ranking. This verifies that the use of the Pareto-based ranking procedure, using the automatically extracted accessibility scores, for the various impairments, leads to rankings which are closer to the perception of the impaired users.

In a further experiment, the multi-objective rankings computed for each session of each user were directly compared to the ground truth rankings using a different comparison measure. For each user session, the ground truth ranking is the one that considers the images in descending order of their ground truth accessibility scores $a_{i,gt}$, so that the first image is the one that the specific user perceived most completely. As a measure to compare between the multi-objective and the ground truth ranking, the Ordered Residual Kernel (ORK) [8, 38] measure was used. The ORK measure computes the similarity between two ranked sets, by counting the common elements in the first positions of both sets, and gradually considering larger sets of top-ranked elements. It takes values in the $[0, 1]$ range, with 0 meaning that the two rankings have no common elements and 1 meaning that the two rankings are identical. The ORK measure was computed for each session of each user. Figure 7 depicts a histogram of the ORK values computed. It can be observed that most ORK values are larger

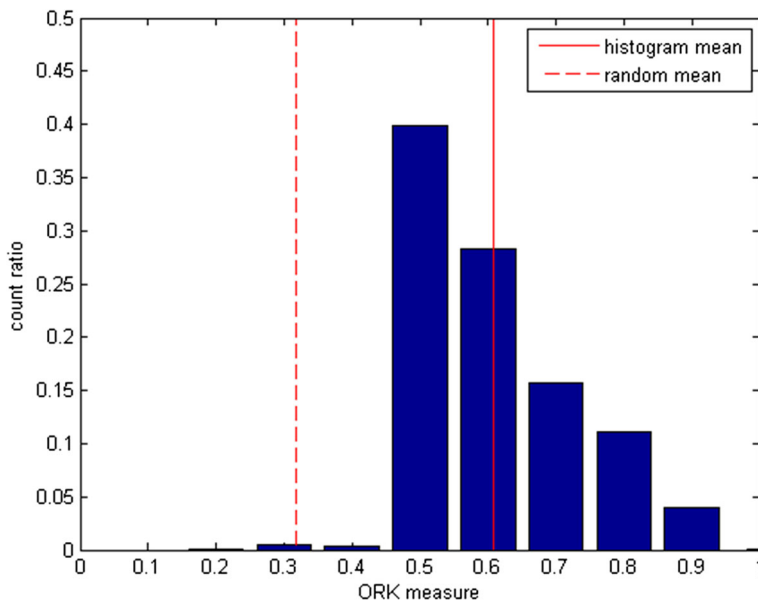


Fig. 7 Histogram of the ORK measure values for all sessions of all users. The *solid red line* denotes the mean ORK value. The *dashed red line* denotes the mean value of 10000 ORK values computed for random rankings. It can be observed that the histogram mean is significantly larger than the random value

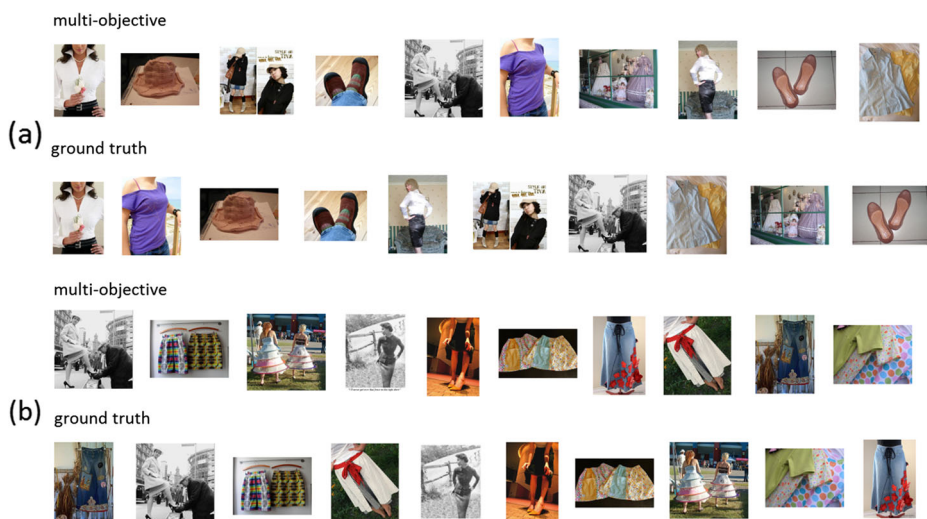


Fig. 8 Example rankings of two user sessions. In each case, the *top row* depicts the ranking produced by the proposed multi-objective method, while the *bottom row* depicts the ground truth ranking. **a** Session of a user with medium cataract and medium glaucoma. **b** Session of a user with high protanopia and mild glaucoma

than 0.5, with a mean of 0.607, denoted in the figure with a solid red line. This indicates that most of the multi-objective rankings were close to the ground truth rankings. As a comparison, the mean of the ORK values for 10000 pairs of random rankings is also depicted in the figure with a dashed red line, at 0.317. The large difference between the histogram mean and the random rankings mean is an indication that the results depicted in the histogram are unlikely to have occurred by chance.

Figure 8 illustrates two further example rankings, for a user with medium cataract and medium glaucoma (a) and a user with high protanopia and mild glaucoma (b). In each example, the top row depicts the multi-objective ranking, produced by the proposed method, while the bottom row depicts the ground truth ranking, based on the ground truth accessibility scores of the images. It can be observed that the automatically produced rankings generally agree with the ground truth rankings, placing at the top positions those images that were easiest for the users to perceive.

7 Conclusion

In this paper, a novel methodology for incorporating accessibility information, in order to rerank the results of a multimedia search engine, was presented. The methodology was applied to an image search engine, in the context of vision-related impairments. The results of the image search engine are reranked so that ones which are easier to see for a user having vision impairments are promoted.

In order to incorporate accessibility information in the reranking procedure, accessibility-related scores were extracted from the images, similar to their relevance scores to a query. The procedure for extracting the accessibility features employed the use of vision filters simulating how an average person having a specific vision impairment would see a result image. The original and the filtered images are compared and a measure of the information

loss caused by the filter is calculated. This measure constitutes the accessibility score of the image. Hereby, five vision impairments were considered, namely cataract, glaucoma, protanopia, deuteranopia and tritanopia.

If only one impairment was considered, ranking the results according to their accessibility scores would suffice to provide an accessibility-based ranking. However, since a user may have more than one impairments, a multi-objective approach was followed in order to calculate a set of rankings, namely the Pareto-optimal ones, which represent different trade-offs among the multiple impairments. The most appropriate one for the user was selected based on the user impairment profile.

The proposed reranking method has been evaluated with two image datasets, using both simulated and real impaired users. Ranking the results according to the proposed method resulted in better correspondence to the perception of impaired users, both qualitatively and quantitatively, confirming that the approach can be successfully applied in order to enhance image search engines by personalizing the results based on the vision impairments of the users. Directions for future work include the extension of the accessibility score extraction procedure to other types of vision or auditory impairments, in order to cover other types of multimedia as well, such as videos or sounds.

Acknowledgments This work was partially supported by the EU funded projects CUBRIK (FP7-287704) and Prosperity4All (FP7-610510).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Adomavicius G, Tuzhilin A (2011) Context-aware recommender systems. In: Recommender systems handbook. Springer, pp 217–253
2. Atrey PK, Anwar HM, El Saddik A, Kankanhalli MS (2010) Multimodal fusion for multimedia analysis: a survey. *Multimed Syst* 16(6):345–379
3. Backstrom L, Kleinberg J, Kumar R, Novak J (2008) Spatial variation in search engine queries. In: Proceedings of the 17th international conference on World Wide Web. ACM, pp 357–366
4. Belkin NJ, Kantor P, Fox EA, Shaw JA (1995) Combining the evidence of multiple query representations for information retrieval. *Inf Process Manag* 31(3):431–448
5. Brettel H, Viénot F, Mollon JD (1997) Computerized simulation of color appearance for dichromats. *JOSA A* 14(10):2647–2655
6. Cao H, Jiang D, Pei J, Qi H, Liao Z, Chen E, Li H (2008) Context-aware query suggestion by mining click-through and session data. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 875–883
7. Chen Y-Y, Suel T, Markowetz A (2006) Efficient query processing in geographic web search engines. In: Proceedings of the 2006 ACM SIGMOD international conference on management of data. ACM, pp 277–288
8. Chin T-J, Wang H, Suter D (2009) The ordered residual kernel for robust motion subspace clustering. In: Advances in neural information processing systems, pp 333–341
9. Coello CAC, Lamont GB, Van Veldhuizen DA (2007) Evolutionary algorithms for solving multi-objective problems, vol 5. Springer
10. CUBRIK project. <http://http://www.cubrikproject.eu/> (2015)
11. Ehrgott M (2005) Multicriteria optimization, vol 2. Springer, Berlin
12. EU project ACCESSIBLE: accessibility assessment simulation environment for new applications design and development, <http://www.accessible-eu.org/>
13. Fairchild MD (2013) Color appearance models. Wiley

14. Fine EM, Rubin GS (1999) Effects of cataract and scotoma on visual acuity: a simulation study. *Optomet Vis Sci* 76(7):468–473
15. Fox EA, Shaw JA (1994) Combination of multiple searches, pp 243–243. NIST SPECIAL PUBLICATION SP
16. Friedman DS, Jampel HD, Lubomski LH, Kempen JH, Quigley H, Congdon N, Levkovitch-Verbin H, Robinson KA, Bass EB (2002) Surgical strategies for coexisting glaucoma and cataract: an evidence-based update. *Ophthalmology* 109(10):1902–1913
17. Giakoumis D, Kaklanis N, Votis K, Tzovaras D (2013) Enabling user interface developers to experience accessibility limitations through visual, hearing, physical and cognitive impairment simulation, pp 1–22. Universal Access in the Information Society
18. Hirvelä H, Koskela P, Laatikainen L (1995) Visual acuity and contrast sensitivity in the elderly. *Acta Ophthalmologica Scandinavica* 73(2):111–115
19. Ji T-L, Sundareshan MK, Roehrig H (1994) Adaptive image contrast enhancement based on human visual properties. *IEEE Trans Med Imag* 13(4):573–586
20. Johnson GM, Fairchild MD (2001) Darwinism of color image difference models. In: Color and imaging conference, vol 2001, pp 108–112. Society for Imaging Science and Technology
21. Kalamaras I, Drosou A, Tzovaras D (2014) Multi-objective optimization for multimodal visualization. *IEEE Trans Multimed* 16(5):1460–1472
22. Kim H, Han K, Yi MY, Cho J, Hong J (2012) Moviemine: personalized movie content search by utilizing user comments. *IEEE Trans Consum Electron* 58(4):1416–1424
23. Kim IY, De Weck O (2005) Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. *Struct Multidiscip Optim* 29(2):149–158
24. Lavery JR, Gibson JM, Shaw DE, Rosenthal AR (1988) Vision and visual acuity in an elderly population. *Ophthalm Physiol Opt* 8(4):390–393
25. Lawrence S (2000) Context in web search. *IEEE Data Eng Bull* 23(3):25–32
26. Leung KW-T, Lee DL, Lee W-C (2013) Pmse: a personalized mobile search engine. *IEEE Trans Knowl Data Eng* 25(4):820–834
27. Liu F, Clement Y, Meng W (2004) Personalized web search for improving retrieval effectiveness. *IEEE Trans Knowl Data Eng* 16(1):28–40
28. Liu J, Li Z, Tang J, Jiang Y, Hanqing L (2014) Personalized geo-specific tag recommendation for photos on social websites. *IEEE Trans Multimed* 16(3):588–600
29. Manning CD, Raghavan P, Hinrich S et al (2008) Introduction to information retrieval, vol 1. Cambridge University Press, Cambridge
30. National eye institute. <https://nei.nih.gov> (2015)
31. Nikulin Y, Miettinen K, Mäkelä MM (2012) A new achievement scalarizing function based on parameterization in multiobjective optimization. *OR Spect* 34(1):69–87
32. Sang J, Changsheng X, Dongyuan L (2012) Learn to personalized image search from the photo sharing websites. *IEEE Trans Multimed* 14(4):963–974
33. Tajima S, Komine K (2015) Saliency-based color accessibility. *IEEE Trans Image Process* 24(3):1115–1126
34. Thang TC, Ro YM (2004) Visual content adaptation for low vision users in mpeg-21 framework. In: International conference on image processing, 2004. ICIP'04. 2004, vol 2, pp 993–996. IEEE
35. Wang M, Sheng Y, Bo L, Hua X-S (2010) In-image accessibility indication. *IEEE Trans Multimed* 12(4):330–336
36. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
37. Web content accessibility guidelines (WCAG). <http://www.w3.org/TR/WCAG20/> (2008)
38. Wong HS, Chin T-J, Jin Y, Suter D (2011) Dynamic and hierarchical multi-structure geometric model fitting. In: 2011 IEEE International conference on computer vision (ICCV), pp 1044–1051. IEEE
39. World health organization. <http://www.who.int> (2015)
40. Xiang B, Jiang D, Pei J, Sun X, Chen E, Li H (2010) Context-aware ranking in web search. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval, pp 451–458. ACM
41. Yang S, Ro YM, Nam J, Hong J, Choi SY, Lee J-H (2004) Improving visual accessibility for color vision deficiency based on mpeg-21. *ETRI J* 26(3):195–202
42. Zhang L, Zhang L, Mou X, Zhang D (2011) Fsim: a feature similarity index for image quality assessment. *IEEE Trans Image Process* 20(8):2378–2386
43. Zitzler E, Laumanns M, Thiele L, Zitzler E, Zitzler E, Thiele L, Thiele L (2001) SPEA2: improving the strength pareto evolutionary algorithm



Dr. Ilias Kalamaras received his diploma in electrical and computer engineering from Aristotle University of Thessaloniki in 2010. He also holds a PhD in multimodal dimensionality reduction from the Intelligent Systems and Networks Group of Imperial College London since 2016. He is a research assistant in the Information Technologies Institute at the Centre for Research and Technology Hellas. His main research interests include retrieval and visualization of multimodal data in multimedia databases.



Dr. Nikolaos Dimitriou received the Diploma degree in Electrical and Computer Engineering from the Aristotle University of Thessaloniki in 2007. From the same institution he received the PhD degree in 2014. As a graduate student he worked as a teaching and research assistant at the Information Processing Laboratory of the Department of Electrical and Computer Engineering. He has also worked as a software developer in the private sector. Since the July of 2014, he works a postdoctoral research associate for the Information Technologies Institute at the Centre for Research and Technology Hellas. His research focuses on the areas of video segmentation, object tracking and recognition.



Dr. Anastasios Drosou is a post-doctoral research fellow at the Information Technologies Institute at the Centre for Research and Technology Hellas. He received his Diploma in Electrical and Computer Engineering, with an expertise in Telecommunications, from Aristotle University of Thessaloniki, and an MSc. in Communication Electronics from the Technische Universität München in 2004 and 2007, respectively. He also holds a PhD in Signal and Image processing from the Imperial College London since 2013. His research interests include biometric security, visualization and network security, regarding the identification. He serves as a regular reviewer for several technical journals and he is the coauthor of more than 20 scientific papers (journals, conferences and book chapters).



Dr. Dimitrios Tzovaras is a senior researcher (Grade A) and also the director of the Information Technologies Institute at the Centre for Research and Technology Hellas. He is also a visiting professor at Imperial College London. Prior to his current position, he was a senior researcher in 3D imaging at Aristotle University of Thessaloniki. His main research interests include virtual reality, assistive technologies, 3D data processing, and stereo and multiview image sequence coding. His involvement with those research areas has led to the coauthoring of more than 60 papers in refereed journals and more than 150 papers in international conferences. He is a member of the IEEE.